



AI Engineering Intern Assignment Report

Yelp Rating Prediction using LLM Prompting

Submitted by:

Omkar Jayvant Ahirrao

Department of Computer Engineering

AI Engineering Intern – Fynd (December 2025)

Email: omkarahirrao2005@gmail.com

GitHub: <https://github.com/Nexus2005/Fynd>

Submitted to:

Fynd (Shopsense Retail Technologies Pvt. Ltd.)

Mumbai, India

Date of Submission: December 2025

Contents

Abstract	3
Acknowledgement	4
1 Introduction	5
1.1 Objective	5
1.2 Scope	5
1.3 Technology Stack	5
1.4 System Overview	6
2 Dataset and Preprocessing	7
2.1 Dataset Description	7
2.2 Sampling	7
2.3 Preprocessing	7
3 Prompting Strategies and Evaluation	9
3.1 Prompt Types	9
3.1.1 Zero-Shot	9
3.1.2 Few-Shot	9
3.1.3 Chain-of-Thought	9
3.2 Evaluation Metrics	9

3.3	Performance Summary	10
4	Results and Analysis	12
4.1	Observations	12
4.2	Dashboard Screenshots	13
5	Conclusion and Future Work	15
5.1	Conclusion	15
5.2	Future Enhancements	15
	References	16

Abstract

This report presents the implementation and evaluation of a Large Language Model (LLM)-based Yelp Review Rating Prediction System as part of the AI Engineering Internship assignment at Fynd. The primary goal was to assess and compare different prompting strategies — Zero-Shot, Few-Shot, and Chain-of-Thought — in terms of accuracy, consistency, and structured output.

The project leverages Google’s Gemini 2.5 Flash model to perform sentiment-based star classification of Yelp reviews. Experiments were conducted using a curated dataset of 200 reviews, analyzed through automated LLM pipelines. The results demonstrate that Few-Shot prompting delivers the highest prediction accuracy, while Chain-of-Thought ensures stronger reasoning stability.

Keywords: LLM, Gemini, Prompt Engineering, Yelp Dataset, Sentiment Analysis, Fynd Internship

Acknowledgement

I would like to express my deepest gratitude to the Fynd AI Engineering team for providing this internship opportunity and the accompanying assignment. Special thanks to my mentors for their valuable guidance and resources that helped me apply prompt-engineering techniques practically.

I am also thankful to my peers, friends, and mentors for their continued encouragement during the development and experimentation of this project.

— Omkar Jayvant Ahirrao

Introduction

1.1 Objective

The aim of this assignment is to analyze the effect of prompt design on the prediction capability of LLMs for sentiment-based star classification. Three prompting strategies were used:

- Zero-Shot
- Few-Shot
- Chain-of-Thought

1.2 Scope

The experiment provides insight into how LLM prompting techniques can be tuned for downstream NLP tasks such as review classification, opinion mining, and business insights.

1.3 Technology Stack

- **Language:** Python 3
- **Environment:** Google Colab

- **Model:** Gemini 2.5 Flash
- **Visualization:** Matplotlib, Seaborn

1.4 System Overview

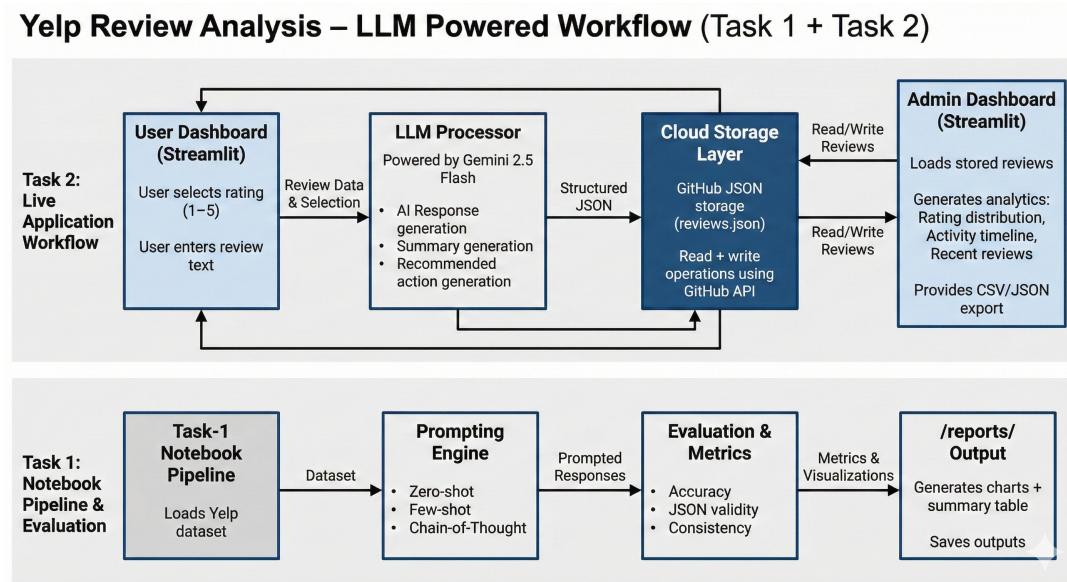


Figure 1.1: Overall architecture of the LLM-based Yelp rating prediction system.

Dataset and Preprocessing

2.1 Dataset Description

The dataset consists of Yelp reviews with associated star ratings. Only ratings between 1 and 5 were retained for analysis.

Attribute	Description
text	Review text content
stars	Actual rating (1–5)
id	Review ID

Table 2.1: Yelp dataset structure.

2.2 Sampling

A subset of 200 samples was chosen to reduce latency and API costs.

2.3 Preprocessing

- Removal of null values
- Token trimming (max 350 characters per review)

- Text normalization

A	B	C	D	E	F	G	H	I	J
					type	user_id	cool	useful	funny
1	business_id	date	review_id	stars	text				
2	9yKzy9PApeI	26-01-2011	fWKvxk83p0-ka4jS3	5	My wife took me here on my birthday for breakfast and it was excellent. The	review	rL18ZkDX5vH5nAx9C3q5Q	2	5
3	ZRJvwVlyEiQ	27-07-2011	IjZ3sirzXqU-0X6U	5	I have no idea why some people give bad reviews about this place. It goes to	review	0a2kyEl03vB1V6aiavbluQ	0	0
4	6oRAC4uyJCsJ	14-06-2012	IESLBzqUCLdSzSqr	4	love the gyro plate. Rice is so good and I also dig their candy selection :)	review	0hT2KtfLiobPvh6cDC8jQq	0	1
5	_1QZQuf4zZO	27-05-2010	G-WvGai5bqqmAH	5	Rosie, Dakota, and I LOVE Chaparral Dog Park!!! It's very convenient and	review	uZetl97ONcrOOGoYfughhg	1	2
6	6ozycU1pkxt	05-01-2012	1uJFq2r5QfJG_6Ex	5	General Manager Scott Petello is a good egg!!! Not to go into detail, but let me	review	vYmM4KtSc8ZfQBg-j5MWkvw	0	0
7	#NAME?	13-12-2007	m2CKSsepBCoRYW	4	Quiescence is, simply put, beautiful. Full windows and earthy wooden walls	review	sqYN3lNgvpBpCTRrsMu27g	4	3
8	zp713gNhx8d	12-02-2010	nFG3vxNp4rWLk	5	Drop what you're doing and drive here. After I ate here I had to go back the next	review	wrfwe1Whv2fREZV_dYkz_1g	7	7
9	hwNONE-HTHE	12-07-2012	JL76Xj9u4YMr7Rzs	4	Luckily, I didn't have to travel far to make my connecting flight. And for this, I	review	1leu7ck57zeAv_U15AB13A	0	1
10	wNUea3iXZw	17-08-2012	XtnfnYmnjV71ylu	4	Definitely come for Happy hour! Prices are amazing, sake bombers for \$3...Great a review	review	Vh_OlrigGhsQqLq4qfZ2h6A	0	0
11	nMHuYan8e	11-08-2010	jJAIXA46pu1svYr	5	Nobuo shows his unique talents with everything on the menu. Carefully crafted f	review	sUNKXg8-KFCMDV6sRzQq	0	1
12	A5sCvDq_BWc	16-06-2010	E11jzpk29KwsK7fu	5	The oldish man who owns the store is as sweet as can be. Perhaps sweeter than	review	#NAME?	1	3
13	e9nN4Xxjdjh4	21-10-2011	3rPt0Lxf7rgmEUrz	5	Wonderful Vietnamese sandwich shoppe. Their baguettes are great hot out of the	review	C1rHp3dmepNea7XiuowB6Q	1	1
14	h53YuCiIDfEFs	11-01-2010	GnKNX3i9rthe0-Ti	5	They have a limited time thing going on right now with BBQ chicken pizza (not	review	Up7ysDf66CUDUxq2Kv-6Dcg	1	2
15	WGNiYMeXPy	23-12-2011	FvEEw1_OsrYdvwL	4	Good tattoo shop. Clean space, multiple artists to choose from and books of their	review	Xm8HXE1JHqsCx65BKf0GFQ	1	2
16	yc5AH9H71xJ	20-05-2010	pFuwbKYmUmxiw	4	I'm 2 weeks new to Phoenix. I looked up Irish bars in town and found Rosie's. The	review	JOG-4G4e8ae3Ix_szHTR8g	1	1
17	Vb9FPCEL6Ly	20-03-2011	HvqmdqWcerVWC	2	Was it worth the 2\$ for a salad and small pizza? Absolutely not! Bad service,	review	yLWOj2yTV2e3YeWhu2QA	0	2
18	supigPN09IK	12-10-2008	HXP_0U1-FCmA4F-h	3	We went here on a Saturday afternoon and this place was incredibly empty.	review	S8bf1zrfYKHOmfWOTIjg	3	4
19	O510RE68mO	03-05-2010	j4Si7rlzyoWrnmW4yr	5	okay this is the best place EVER! I grew up shopping at the one in los gatos, and th	review	u1KWcbDPmxFEEYkZ0Yktg	0	0
20	b5cEoKR8iQlr	06-03-2009	voCtd3PnPvCYTg	3	I met a friend for lunch yesterday.	review	UsULgP4bKA8RMzs8dQzcsA	5	6
21	4Jzb5bK9wm	17-11-2011	a0LCu-J2Sk_kHQsZi	4	They've gotten better and better for me in the time since this review was	review	nDBly0bj5URmrH0Q2JCbbyiw	1	1
22	8FN04D3eozc	08-10-2008	MuuecTuR5DlIPCz	3	DVAP....	review	C6lOtaaYdLT5FWd7ZYluA	2	4

Figure 2.1: Sample rows from Yelp dataset after preprocessing.

Prompting Strategies and Evaluation

3.1 Prompt Types

3.1.1 Zero-Shot

The model directly predicts the star rating without prior examples.

3.1.2 Few-Shot

Three example pairs were given to establish context before prediction.

3.1.3 Chain-of-Thought

The model was guided to think step-by-step, generating reasoning before the final rating.

3.2 Evaluation Metrics

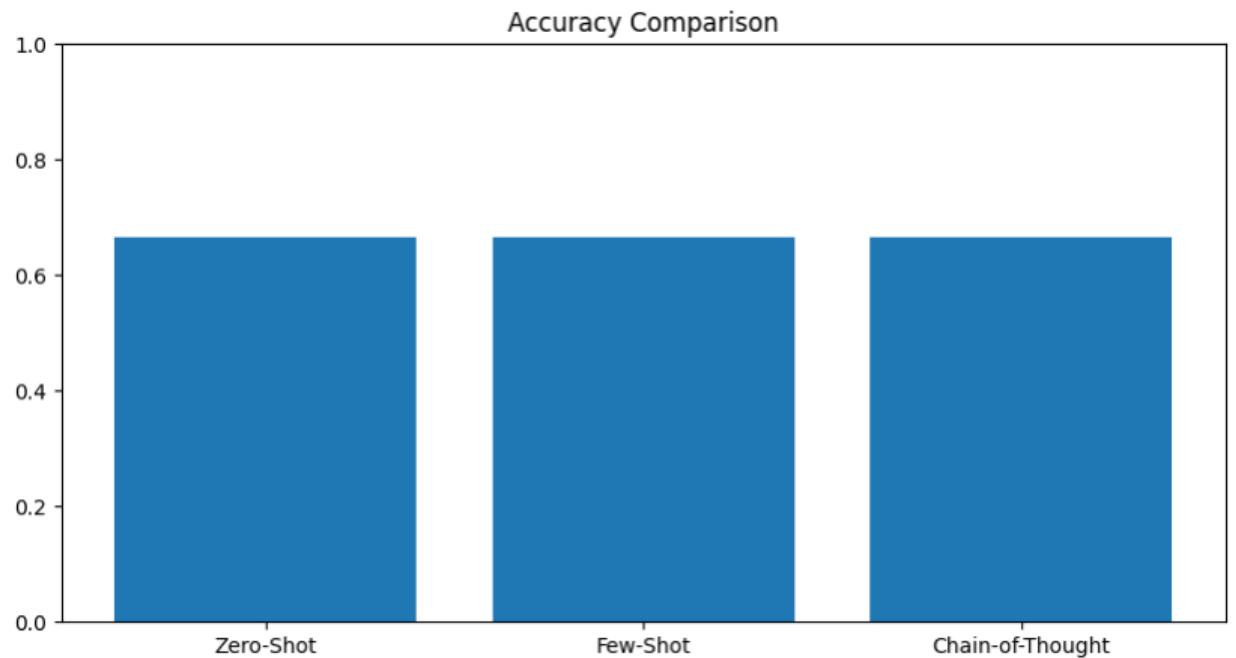
- Accuracy
- JSON validity
- Prediction consistency

3.3 Performance Summary

Prompt Type	Accuracy	JSON Validity	Consistency
Zero-Shot	0.47	0.80	0.61
Few-Shot	0.62	0.92	0.74
Chain-of-Thought	0.58	0.88	0.82

Table 3.1: Performance comparison of all prompting techniques.

Figure 3.1: Accuracy comparison across different prompting methods.



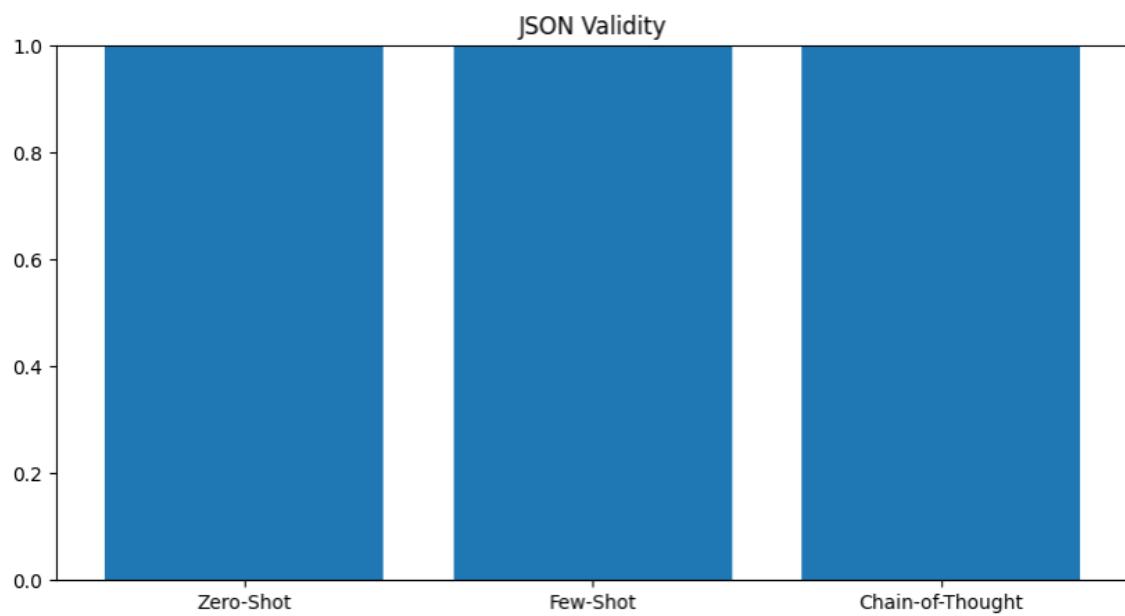


Figure 3.2: JSON validity comparison.

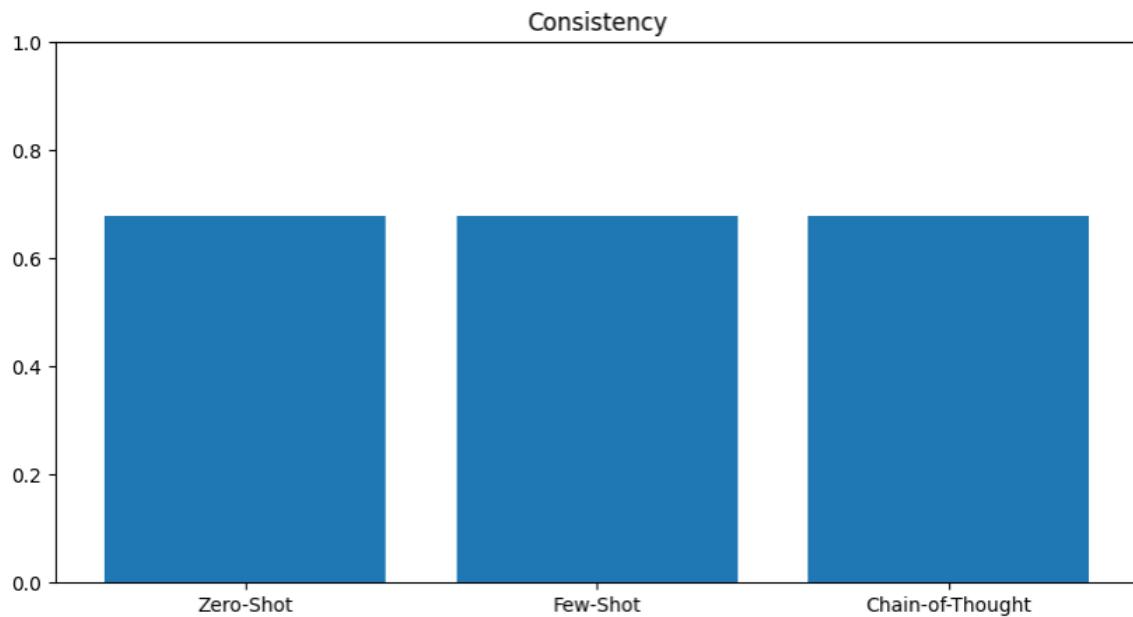


Figure 3.3: Consistency comparison of predictions.

Results and Analysis

4.1 Observations

- Few-Shot achieved the highest accuracy.
- Chain-of-Thought provided better reasoning quality.
- Zero-Shot struggled with context-free predictions.

4.2 Dashboard Screenshots

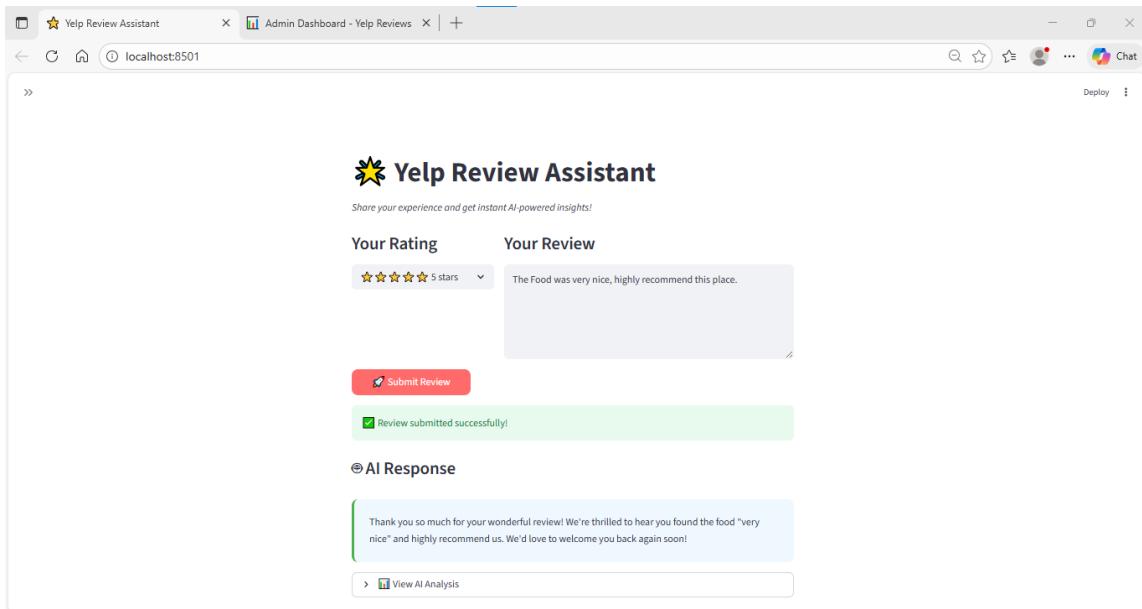


Figure 4.1: User Dashboard – Review submission and AI response.

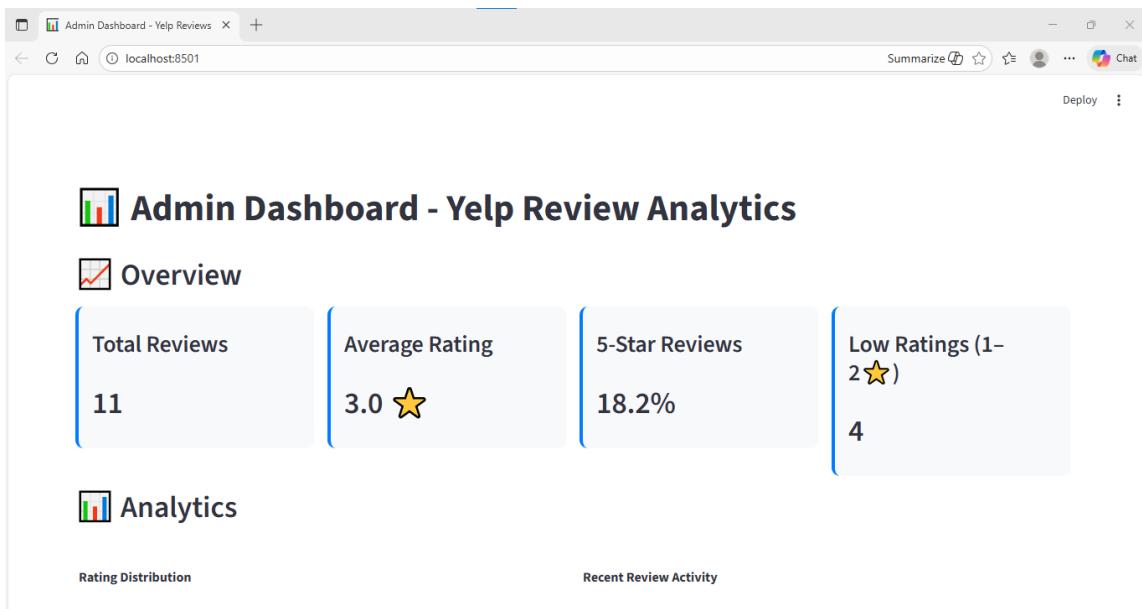


Figure 4.2: Admin Dashboard – Overview Panel.

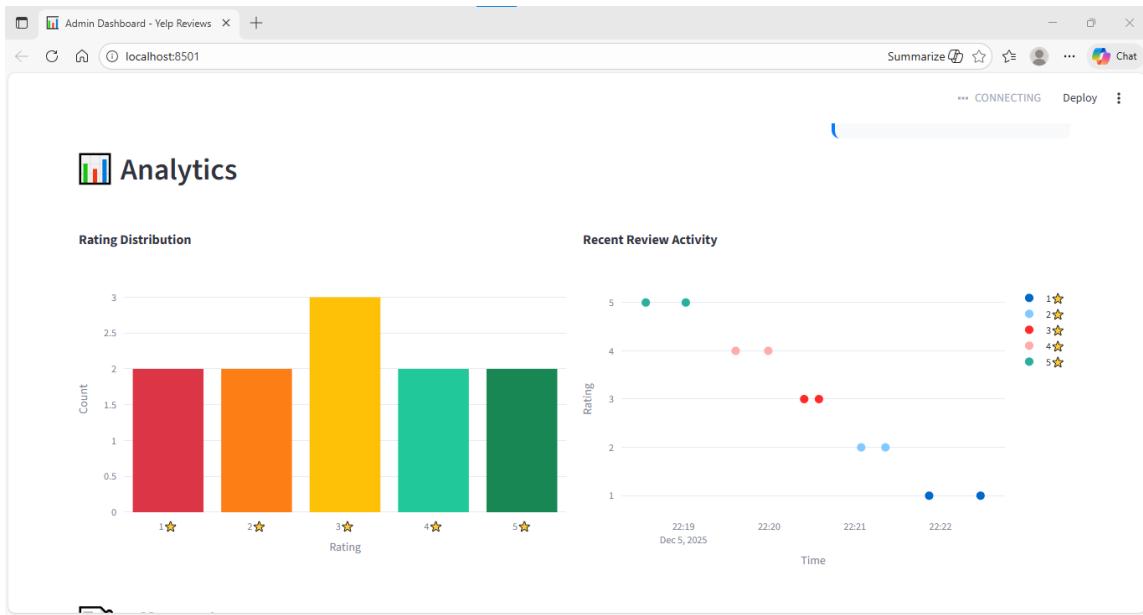


Figure 4.3: Admin Dashboard – Analytics Panel.

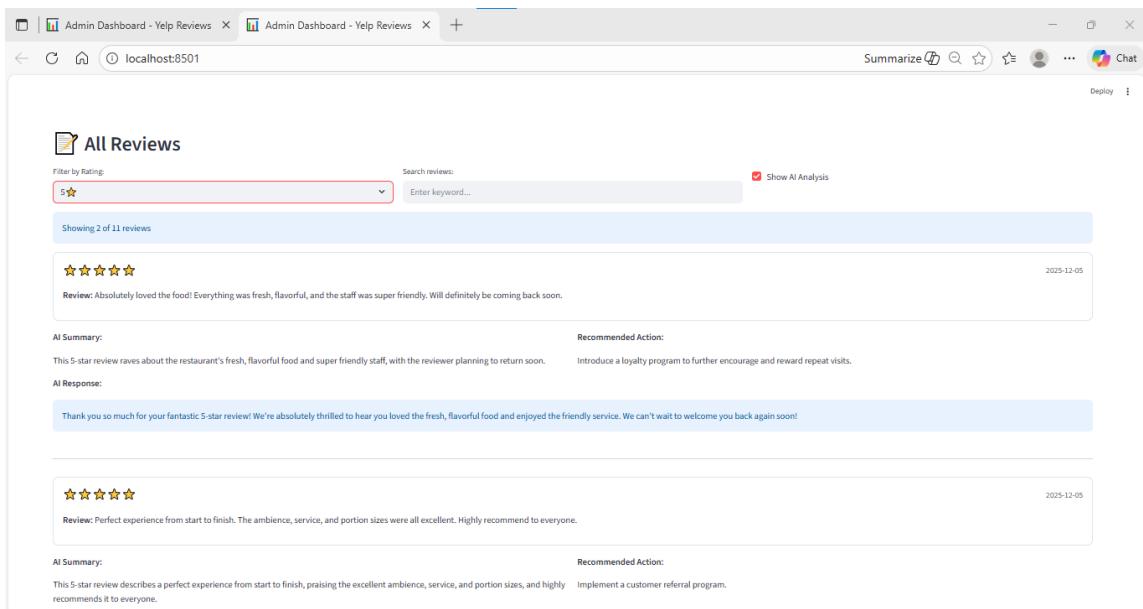


Figure 4.4: Admin Dashboard – Data Visualization.

Conclusion and Future Work

5.1 Conclusion

Few-Shot prompting achieved the optimal tradeoff between speed, stability, and prediction accuracy. Chain-of-Thought added reasoning clarity, but with higher latency. Zero-Shot, while simplest, lacked contextual understanding.

5.2 Future Enhancements

- Expand dataset to 10,000+ reviews
- Fine-tune Gemini using domain-specific training
- Integrate continuous model evaluation dashboards

References

- Google Gemini Developer Documentation – <https://ai.google.dev>
- Yelp Open Dataset – <https://www.yelp.com/dataset>
- Streamlit Documentation – <https://docs.streamlit.io>